DOCUMENT RESUME

ED 475 143 TM 033 736

AUTHOR	Bong, Mimi; Hocevar, Dennis
TITLE	Measuring Self-Efficacy: Multi-Trait Multi-Method Comparison of Scaling Procedures.
PUB DATE	2001-04-00
NOTE	31p.; Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA, April 10-14, 2001).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Concurrent Validity; Foreign Countries; *High School Students; High Schools; *Self Efficacy
IDENTIFIERS	Confirmatory Factor Analysis; *Discriminant Validity; Motivated Strategies for Learning Questionnaire; South Korea

ABSTRACT

Convergent and discriminant validity of various self-efficacy measures was examined across two studies. In Study 1, U.S. high school students (n=358) rated their self-efficacy in 6 school subjects with reference to specific problems or general self-efficacy statements on the Motivated Strategies for Learning Questionnaire (MSLQ) (P. Pintrich and E. De Groot, 1990). In Study 2, Korean female high school students (n=235) judged their perceived efficacy with reference to specific problems, specific task descriptions, and MSLQ statements in 3 school subjects. Across both studies, the first-order confirmatory factor analyses provide support for both convergent validity of different self-efficacy responses and discriminant validity of perceived self-efficacy across different subject areas. The second-order confirmatory factor analyses confirmed the discriminant validity of self-efficacy beliefs. Substantial method effects were also observed. The problem- and task-referencing methods correlated with each other to a greater extent than they did with the MSLQ self-efficacy scale. (Contains 1 figure, 6 tables, and 45 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 475 143

Measuring Self-Efficacy: Multi-Trait Multi-Method Comparison of Scaling Procedures

Mimi Bong
University of South Carolina

Dennis Hocevar
University of Southern California

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

M. Bong

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract

Convergent and discriminant validity of various self-efficacy measures was examined across two studies. In Study 1, US high school students (N = 358) rated their self-efficacy in six school subjects in reference to either specific problems or general self-efficacy statements on the Motivated Strategies for Learning Questionnaire (MSLQ). In Study 2, Korean female high school students (N = 235) judged their perceived efficacy in reference to specific problems, specific task descriptions, and MSLQ statements in three school subjects. Across Studies 1 and 2, the 1st-order CFAs provide support for both convergent validity of different self-efficacy responses and discriminant validity of perceived self-efficacy across different subject areas. The 2nd-order CFAs confirmed the discriminant validity of self-efficacy beliefs. Substantial method effects were also observed. The problem- and task-referencing methods correlated with each other to a greater extent than they did with the MSLQ self-efficacy scale.

Mimi Bong, Department of Educational Psychology, University of South Carolina.
Dennis Hocevar, Division of Learning and Instruction, University Southern California.

In J. Baumert (Chair), *Multiple frames of reference and self-related cognitions: An international perspective.* Symposium conducted at the annual meeting of the American Educational Research Association, Seattle, WA, April 2001.

TM033736

BEST COPY AVAILABLE

THIS PAGE INTENTIONALLY LEFT BLANK

The primary purpose of the present investigation was to assess the equivalence of self-efficacy judgments that were measured by different methods. Convergent and discriminant validity of academic self-efficacy responses were examined in a multi-trait multi-method (MTMM) framework. In Study 1, US high school students reported their self-efficacy perceptions in six school subjects by rating either their confidence for solving specific problems presented or their agreement with each of the self-efficacy statements provided. In Study 2, Korean high school students reported their self-efficacy in three school subjects in reference to either specific problems, written task descriptions, or general self-efficacy statements. Confirmatory factor analyses (CFA) and higher-order confirmatory factor analyses (HCFA) were applied to these MTMM self-efficacy data.

Brief Overview of Self-Efficacy Research

Self-efficacy refers to one's convictions to successfully organize and execute a course of action that is required to achieve a desirable outcome (Bandura, 1997). It is context-specific judgment that is closely tied to the specific domain and situation in question (Zimmerman, 1995). Academic self-efficacy, in particular, represents learners' subjective confidence for successfully performing given academic tasks at designated levels (Schunk, 1991). As such, it wields a critical influence on virtually all aspects of student learning. Students with a strong sense of self-efficacy willingly choose challenging academic tasks (Bandura & Schunk, 1981), use effective learning strategies (Pintrich & De Groot, 1990), persist longer in the face of difficulties (Lent, Brown, & Larkin, 1984), and set higher academic goals (Zimmerman, Bandura, & Martinez-Pons, 1992). These students also demonstrate more positive attitudes and emotions toward learning as evidenced by their higher academic aspirations and lower depression (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996), lower anxiety (Pajares & Miller, 1994), and lower apprehension (Pajares, Miller, & Johnson, 1999) in academic contexts. Through its positive influence on subsequent motivation and learning, heightened self-efficacy brings about better academic performance (see Multon, Brown, & Lent, 1991).

As evidence demonstrating the potency of academic self-efficacy beliefs accumulates, an increasing number of researchers have incorporated this important construct in their investigations. Up until now, operational definitions of self-efficacy have been relatively more consistent compared to those of other self-constructs (Bong & Clark, 1999). Still, they are not without some discrepancy. Investigators have used different methods to assess self-efficacy perception, which sometimes renders comparability of the findings unclear. In part, the problem resonates Pajares' (1996) comment on the specificity of self-efficacy beliefs and their correspondence to criterial tasks. The outcome of interest in educational research ranges from performance on a very specific task to more general-level indicators such as course choice or semester grades. Because outcomes like course grades typically reflect some form of aggregation of students' performances on diverse tasks and activities, they pose additional complexity to self-efficacy assessment. Given the growing trend in self-efficacy research, it is important to evaluate different assessment methods in relation to the self-efficacy theory.

Measuring Self-Efficacy

There have been four broad categories of measurement techniques that researchers use to assess the strength of self-efficacy beliefs. The first and standard method of measuring academic self-efficacy is to present a set of specific problems, performance on which is the very target of

prediction. Students report their confidence for successfully solving each type of problems on a 0 to 100 scale with a 10-unit interval. The following verbal descriptors usually accompany the scale: 10 (not sure), 40 (somewhat sure), 70 (pretty sure), and 100 (very sure). Schunk and his colleagues repeatedly used this method for measuring elementary school students' arithmetic self-efficacy (Schunk, 1982, 1983; Schunk & Cox, 1986; Schunk & Gunn, 1986; Schunk & Hanson, 1985, 1989; Schunk, Hanson, & Cox, 1987). Zimmerman and Martinez-Pons (1990), in their comparison of gifted- and regular-school students' self-efficacy perceptions, presented verbal (i.e., word defining) and math problems of increasing difficulty (i.e., simple arithmetic, algebra, probability, and statistics) and asked students to rate their perceived capability to solve each of the problems. Zimmerman and Kitsantas (1999) likewise obtained students' self-efficacy ratings for a sentence-combining task by presenting specific writing revision problems.

The second category of self-efficacy assessment method is similar to the first category in that it provides concrete anchors which respondents use for gauging their efficacy perceptions. What are being presented are not specific problems but verbal descriptions of specific task components that reflect the major aspects of successful performance. Researchers choose this method when the target performance cannot easily be summed up as specific problems. In reading, for example, students are asked to judge their confidence to successfully perform tasks such as: Read one of the textbooks, know all the words on a page in one of the schoolbooks, know the meaning of plurals, prefixes, and suffixes, and understand the main idea of a story (Shell, Colvin, & Bruning, 1995). In writing, students estimate their confidence for performing such tasks as: Write a one-page summary of a book, correctly punctuate a sentence, correctly spell all words in a one-page story or composition, and correctly use parts of speech such as nouns, verbs, adjectives, or adverbs (Pajares et al., 1999; Shell et al., 1995). This method is also used often to describe computer-related skills: e.g., Use Hypercard clip art, create a background design that is used by multiple cards, download necessary materials from the Web, use Internet search engines such as Yahoo (Joo, Bong, & Choi, 2000; Schunk & Ertmer, 1999).

Whereas these first two methods concentrate on specific facets of task performance, the latter two methods concentrate more on the overall performance levels. One of them is to ask students about their confidence to achieve a specific letter grade. Students rate the strength of their beliefs that they could obtain each of the letter grades ranging from A to F (Zimmerman & Bandura, 1994). The other method is to ask students to judge their general confidence to function successfully in the given domain without making an explicit reference to any individual problems or tasks. Instead, descriptions of generic tasks that are commonly performed in most academic domains are provided in the context of specific subjects. Therefore, students rate how much they agree with statements like: I am certain that I can understand what is taught in (a specific subject) class, I expect to do very well in (a specific subject) class, and I am certain that I can figure out how to do the most difficult schoolwork in (a specific subject) (Pintrich & De Groot, 1990).

To date, several researchers have addressed the issue of self-efficacy scale differences in terms of predictive validity. For example, Pajares and Miller (1995) presented convincing evidence that how one assesses self-efficacy judgment could produce different results regarding its relationships with relevant outcomes. The researchers solicited college students' confidence ratings for either solving specific math problems, completing everyday math tasks, or performing successfully in math-related courses. As expected, math problems self-efficacy was a better predictor of math problem-solving performance than math courses self-efficacy. Math courses self-efficacy predicted choice of math-related majors better than math problems self-efficacy. The three self-efficacy scores were all highly correlated among themselves as well as with the two outcome measures. Bong (in press-b) also compared multiple self-efficacy scores that were

assessed at varying levels of specificity. College students reported their confidence for correctly solving specific problems presented, successfully mastering the representative topics of the course, successfully performing in the course, and performing well in college courses in general. As Pajares and Miller observed, all self-efficacy scores were positively correlated among themselves and, with an exception of problem-specific self-efficacy, with the value students perceived in the course. More interesting, correlation between any two self-efficacy scores decreased as the difference in their measurement levels increased.

These investigations are instrumental in establishing the basic guidelines for assessing self-efficacy beliefs. The positive correlation among different self-efficacy scales reported in both studies also provides some evidence of convergent validity. However, issues of convergent and discriminant validity of scales can be dealt with more effectively in a multi-trait multi-method design (Campbell & Fiske, 1959), which requires a minimum of two traits[1] assessed by at least two methods. Because most previous studies measured self-efficacy perceptions in relation to a single domain, convergent and discriminant validity of self-efficacy responses have not been probed systematically according to the Campbell and Fiske criteria. In the present research, multiple measures of self-efficacy beliefs in multiple academic domains were available across two studies, allowing MTMM comparison of self-efficacy scores. Unlike traditional MTMM analysis, this study applied confirmatory factor analysis and higher-order confirmatory factor analysis. CFA is especially useful in situations where linkages between observed variables and latent constructs can be clearly established according to the theory. Because most measures in social and behavioral sciences contain sizable measurement errors, CFA affords important advantages over a zero-order correlation approach. Rather than relying on an unrealistic assumption that the measures are perfect, CFA takes the measurement errors into account. In the MTMM context, it also allows partitioning of the indicator variance into trait, method, and random error components. Researchers generally agree that CFA is the most defensible and informative approach to the analysis of the MTMM (Marsh, 1993; Marsh & Hocevar, 1983). By applying CFA and HCFA procedures to MTMM self-efficacy matrices, this study aimed at examining (1) the equivalence of self-efficacy responses from different assessment methods (i.e., convergent validity) and (2) the distinctiveness of self-efficacy beliefs in different academic domains (i.e., discriminant validity).

Study 1

Method

Participants

The sample consisted of 358 students (49% boys) enrolled in four high schools in Los Angeles county at the time of the survey. Among the 588 students who participated in the larger research project (see Bong, 1997b, for a description of the larger sample), students who reported having previous experience with all six subject areas were selected. Ethnic composition of the present sample was: 16% White, 6% African American, 55% Hispanic, 21% Asian, and 2% Native American and other. Students were mostly in Grades 11 (21%) and 12 (78%).

Measures and Procedures

Problem-referenced self-efficacy. Seven typical problems from six school subjects (i.e., English, Spanish, American history, algebra, geometry, and chemistry) were prepared from the Scholastic Aptitude Test (SAT) I and II preparatory booklets (Brownstein, Weiner, & Green,

1994; College Entrance Examination Board and Educational Testing Service, 1994; see Bong, 1997b, for sample problems). Care was taken to ensure that problems of representative types and moderate difficulty were included. Each problem was presented through an overhead projector for a duration that was long enough to recognize its type but too short to attempt its solution. Students rated how confident they were to correctly solve the types of problems presented on a scale ranging from 0 to 100 in 10-unit intervals. The following verbal descriptors were provided to help students understand more clearly what each number represented: 0 (not sure), 40 (maybe), 70 (pretty sure), and 100 (real sure). This is a standard procedure of assessing self-efficacy beliefs using specific problems (see, e.g., Bandura, 1997, pp. 42 - 46). One might argue that the difference in response scales (e.g., using a 0 – 100 scale versus using a 1 – 7 scale; see below) could confound the results regarding the different types of measurement. Although this certainly is a possibility, we felt that using the most typical assessment strategies associated with each method would provide more insights as to the difference in measurement methods as being used in the current literature.

MSLQ self-efficacy. Students responded to self-efficacy items on the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich & De Groot, 1990). The MSLQ self-efficacy items seek students' endorsement ratings on statements describing general academic events in the context of specific domains. Of the nine items on the original scale, three ask students to compare their capability to that of their peers. Self-efficacy researchers maintain that judgments of self-efficacy depend more heavily on the mastery criteria (i.e., being able to succeed) than on the normative ones (i.e., being better than others) (Bong & Clark, 1999; Zimmerman 1995, 1996). Accordingly, comparative items were excluded from the current investigation. The final scale contained the following six items for each school subject: "I'm certain that I can understand what is taught in (a specific school subject) class," "I expect to do very well in (a subject) class," "I am sure that I can do an excellent job on the problems and tasks assigned for (a subject) class," "I know that I will be able to learn the material for (a subject) class," "My study skills are excellent in (a subject) class," and "I think I will receive a good grade in (subject) class." Response categories ranged from 1 (not at all true) to 7 (very true) as in Pintrich and De Groot (1990).

Results

Table 1 presents descriptive statistics of the scales. Eighteen measured variables (MVs) were created for Problems Self-Efficacy by combining responses to two to three problems. Specifically, responses to Problems 1, 4, and 7, Problems 2 and 5, and Problems 3 and 6 in each subject were averaged to produce three MVs for each of the six school subjects. Another eighteen MVs were created for MSLQ Self-Efficacy by combining responses to Items 1 and 4, Items 2 and 5, and Items 3 and 6 in each domain (descriptive statistics of individual items and MVs for both scales are available from the first author). Therefore, there were six MVs in each school subject, three of which shared the same method. In both Studies 1 and 2, we decided to use item parcels rather than individual items as indicators for several reasons. First, the sample size of $N = 358$ does not permit such elaborate analyses of using single-item indicators. We acknowledge that the probability of obtaining proper solutions improves with a larger number of indicators per factor when sample size is small and when individual items are used as indicators (Marsh, Hau, Balla, & Grayson, 1998). When using three or more item parcels, however, solutions almost always converge and are unaffected by the sample size. Second, all the scales appear extremely homogeneous. Most reliability coefficients of scales based on individual items (as opposed to

item parcels) exceeded .90. Although homogeneity does not always guarantee unidimensionality of scales, it nonetheless gives us some assurance that the results would have been very similar had individual items been used as MVs in place of item parcels. Finally, item parcels are known to meet the multivariate normality assumption that underlies structural equation modeling better than individual items (Kline, 1998, p. 237).

Insert Table 1 about here

The problem-referencing and the MSLQ were treated as two methods, whereas self-efficacy perceptions in the six school subjects were treated as six traits. The pattern of covariation among these MVs was likely created by many factors, most notably by the method and trait effects. Depending on the relative contribution of each source, the number of factors required to obtain satisfactory model fit would differ. Different CFA models were thus specified and compared. All CFAs were conducted with the EQS program (Bentler, 1992). Because the two methods used very different response scales, we conditioned the Problems Self-Efficacy matrix by dividing all responses by 10.

If observed variation among MVs was mostly due to method effects (i.e., students provided similar ratings to self-efficacy items using the same method regardless of the content domain being tapped), two correlated method factors alone should be able to illustrate the data to a sufficient degree (Model 1). On the other hand, if the data pattern was created mostly because of the different traits (i.e., students' self-efficacy ratings were primarily determined by their self-efficacy beliefs in the subject domain, irrespective of the assessment tools), six correlated trait factors should suffice (Model 2). The nonnormed fit index (NNFI) and comparative fit index (CFI) reported in Table 2 represent roughly the percentage of variance in the data that is accounted for by a given model. Values greater than .90 are commonly taken as evidence of satisfactory model fit. Neither Model 1 (NNFI = .422, CFI = .456) nor Model 2 (NNFI = .671, CFI = .697) was able to reproduce the observed data to a satisfactory degree.

Insert Table 2 about here

Model 3 specified six correlated traits and two correlated method factors. All fit indexes improved substantially, falling only little short of the recommended cut-off value of .90 (NNFI = .872, CFI = .890; see Table 2). This strongly attests to the need for both trait and method factors. Unfortunately, there is inherent danger of partial underidentification when one is dealing with only two methods (Marsh & Hocevar, 1983). Model 4 avoids this problem by specifying twelve first-order factors, each of which represents a unique combination of a particular trait and method (e.g., English self-efficacy assessed by the problem-referencing method). It also allows examining the validity issue according to the Campbell-Fiske guidelines. Model fit improved substantially with fit indexes well above the acceptable value (NNFI = .954, CFI = .961). Each factor was clearly defined with sizable loadings ranging from .776 to .961 (Mdn = .906). Table 3 presents correlation coefficients among the twelve factors.

Insert Table 3 about here

Factor correlation coefficients among the six Problems Self-efficacy factors ranged from .085 to .881, whereas those among the MSLQ Self-Efficacy factors ranged between .127 and .837. Within each method, the highest correlation existed between Algebra and Geometry

Self-Efficacy factors. Because these two factors were very highly correlated, three additional CFA models were run to test whether they could be combined into a single Math Self-Efficacy factor. Model 5 specified 11 factors by combining Problems Algebra and Problems Geometry factors of Model 4 into a single Problems Math Self-Efficacy factor. Model 6 likewise hypothesized 11 factors by combining MSLQ Algebra and MSLQ Geometry factors into an MSLQ Math factor. Model 7 combined Algebra and Geometry Self-Efficacy factors of both Problems and MSLQ methods into Problems Math and MSLQ Math factors, thus specifying only 10 trait-method combination factors. As can be seen in Table 2, in all three instances, combining algebra and geometry indicators to load on the same Math factor resulted in poorer overall model fit compared to Model 4. Loadings of relevant indicators also showed a uniform decline when only a single Math Self-Efficacy factor was specified instead of separate Algebra and Geometry Self-Efficacy factors. Correlation coefficients among other self-efficacy factors within each method were substantially less than 1.0. These results demonstrate discriminant validity of the six subject-specific self-efficacy factors as assessed by each method.

In the first-order CFA, evidence of convergent validity can be found when (1) statistically significant and substantial loadings on the trait factors are obtained and (2) significant decrement in fit is observed when trait factors are deleted from model specification (Gardner, Cummings, Dunham, & Pierce, 1998; Marsh & Hocevar, 1988). Results from Model 4 met both of these basic requirements. Campbell and Fiske (1959) also suggested that convergent validity requires that mono-trait hetero-method correlation coefficients be significant and substantial in magnitude and be higher than hetero-trait mono-method (i.e., method effects) or hetero-trait hetero-method coefficients. Because correlations among CFA factors essentially represent correlations among scale scores corrected for attenuation, the Campbell-Fiske criteria of determining the convergent and discriminant validity can be readily and more accurately applied (Marsh & Hocevar, 1988). As Table 3 reports, the convergent validity (i.e., mono-trait hetero-method) coefficients were clearly higher than the hetero-trait hetero-method coefficients. Convergent validity coefficients were not always higher than the hetero-trait mono-method correlation coefficients in the same column or row, because some of the self-efficacy factors were highly correlated when they shared the same method. Still, on the whole, the mono-trait hetero-method correlation coefficients were generally higher (average $\underline{r}$ = .611) than either the average correlation among Problems Self-Efficacy factors (average $\underline{r}$ = .449) or the average correlation among the MSLQ Self-Efficacy factors (average $\underline{r}$ = .399). Self-efficacy factors for the six subjects also showed a pretty consistent pattern of interrelatedness. Across the two methods, English and History Self-Efficacy and the three math-related self-efficacy factors (i.e., Algebra, Geometry, and Chemistry Self-Efficacy) demonstrated particularly strong correlation.

Although these results are strongly suggestive of convergent validity, the nature of analyses does not permit us to generate precise answers. In particular, it is difficult to separate out the relative contribution of trait and method effects from these trait-method combination first-order factors. Marsh and Hocevar (1988) demonstrated that this could be achieved by applying HCFAs to the MTMM matrix. However, to obtain a fully identified second-order factor structure, a minimum of three trait and three method first-order factors need to be defined.

Study 2

Because only two methods were used in Study 1, it was not possible to analyze relations among trait factors after controlling for the methods effects. Nor was it possible to examine equivalence of different methods after the trait effects had been accounted for. In Study 2, three

different methods were used for assessing self-efficacy judgments across three different subject areas. More specifically, students' self-efficacy perceptions in Korean, English, and math (hence three traits) were assessed by problem-referencing, task-referencing, and MSLQ (i.e., subject-referencing) methods (hence three methods). Therefore, it became possible to examine the aforementioned issues directly by specifying CFA and HCFA models, in addition to interpreting findings more clearly according to the Campbell-Fiske guidelines.

Method

Participants

Participants were 235 students from a Korean female high school. Students completed self-efficacy surveys as part of a research project comparing the predictive utility of different self-efficacy beliefs for immediate and delayed academic performances (Bong, 2001).

Measures and Procedures

Problem-referenced self-efficacy. Problems used for assessing self-efficacy perceptions came from placement tests developed by one of the educational testing services in Seoul, Korea. These tests contained problems of representative types and topics that entering high school students should be able to handle. There were 25 problems in each school subject. In the Korean test, several long reading passages were used, each of which was often referred to by multiple problems. Some of these questions also referred to each other, making it difficult to separate them. In these instances, presenting individual problems was not deemed appropriate because doing so would remove them from their very contexts. When these problems were inspected across passages, several common problem types were identified (e.g., vocabulary, grammar, reading comprehension, etc.). Therefore, problems were reorganized according to their types and each reading passage was presented with one or more problems that it most logically related to. This resulted in reduction in the number of problems presented in Korean (n = 15). Students rated their confidence toward solving problems of the given type when these problems were presented for a brief duration. Procedures were the same as those used in Study 1.

Task-referenced self-efficacy. Ten representative task descriptions in each domain were developed out of the placement test problems. All specific information such as numbers, figures, vocabulary, or reading passages were removed from problems. These problem descriptions were then revised so that they illustrated generic portraits of tasks typically performed in each domain. For example, students read task descriptions such as "Read a given passage and determine its main theme" and "Change given sentences from active to passive voice" in Korean, "Read a given paragraph and fill in parentheses with appropriate conjunctions" and "Find parts that are grammatically incorrect from given sentences" in English, and "Solve equations containing square roots" and "Solve for x in a quadratic equation" in math. A response scale ranging from 0 to 100 was used again with the same verbal descriptors.

MSLQ self-efficacy. Among the six MSLQ items used in Study 1, only five items were retained in Study 2. The item that read "My study skills are excellent in (a subject) class" was dropped. Self-efficacy refers to personal convictions and expectations and items measuring self-efficacy should thus ask whether one is confident that she "can" or "will be able to" execute certain behaviors required for desired outcomes. In this sense, the particular item in question did not exactly appear to tap perceived efficacy. Although empirical results from Study 1 showed that this item behaved similarly to other items, it was removed from the scale for this conceptual reason. A response scale of 1 to 5 was used as in Study 1.

Results

A total of thirty-nine MVs were created (descriptive statistics of individual items and MVs are available from the first author). With English and Math Problems Self-Efficacy, responses to Problems 1, 6, 11, 16, and 21 were aggregated to produce the first MV. Following the next sequences produced another four MVs. In Korean, combining three responses according to the same sequence (starting with Problems 1, 6, and 11) produced five MVs. With Tasks Self-Efficacy, five MVs in each subject were prepared by averaging two responses (starting with Tasks 1 and 6), following similar sequences used in Problems Self-Efficacy. MSLQ responses to Items 1 and 5 and Items 2 and 4 were combined and Item 3 functioned as a single-item indicator. In total, five Problems Self-Efficacy MVs, five Tasks Self-Efficacy MVs, and three MSLQ Self-Efficacy MVs in each school subject were constructed.

Table 4 presents goodness-of-fit indexes for all CFA and HCFA models. Models 1 to 4 in Study 2 shared the same theoretical structure with Models 1 to 4 in Study 1. The only difference between the two studies with regard to these four models was the number of trait and method factors specified. Results from these first-order CFAs were consistent with those obtained in Study 1. Specifying method factors only in the absence of trait effects (Model 1) or trait factors only without method effects (Model 2) resulted in poor model fit, although the trait-only model was somewhat superior compared to the method-only model as was the case in Study 1. Model 3 that separated out trait and method variance at the indicator level by specifying three method and three trait first-order factors also did not reach acceptable fit criteria (NNFI = .879, CFI = .892). Model 4 with nine first-order factors, each of which reflected unique combination of a single method and a single trait, demonstrated the best and satisfactory fit to the empirical data (NNFI = .925, CFI = .932). Factor loadings ranged from .741 to .972 (Mdn = .891).

Insert Table 4 about here

Examining correlation coefficients among these nine factors again permitted applying the Campbell-Fiske rules for determining convergent and discriminant validity. As Table 5 shows, factor intercorrelation in Model 4 provides evidence of convergent validity of traits assessed by different methods as well as discriminant validity of traits assessed by the same method. More specifically, mono-trait hetero-method coefficients between Problems and Tasks Self-Efficacy factors were .680 in Korean, .790 in English, and .712 in math with an average correlation of .727. The three correlation coefficients were higher than all hetero-trait mono-method and hetero-trait hetero-method coefficients in the same column or row. Convergent validity coefficients between Tasks and MSLQ Self-Efficacy factors come in next in magnitude, ranging from .580 to .742 (average r = .656). Those between Problems and MSLQ Self-Efficacy factors were .599 in Korean, .588 in English, and .555 in math, with an average correlation of .581. These coefficients were clearly larger than the hetero-trait hetero-method coefficients. With few exceptions, the mono-trait hetero-method correlation coefficients were also larger than hetero-trait mono-method coefficients. The difference, however, was not definitive. Campbell and Fiske (1959) argued that such a case represents true trait correlation, strong method effects, or both. All correlation coefficients among different self-efficacy factors that shared the same method were well below 1.0 across the three methods, attesting to the discriminant validity of the three subject-specific self-efficacy factors. The Korean, English, and Math Self-Efficacy factors demonstrated an average correlation of .560, .625, and .396 when assessed with Problems, Tasks, and MSLQ, respectively.

Insert Table 5 about here

The fundamental difference of the MTMM matrix in Study 2 from that in Study 1 was the provision of three methods. Whereas trait and method effects had to be directly inferred from the relations among MVs in Study 1, these effects could now be estimated from relations among the nine trait-method combination first-order factors. Second-order trait and second-order method factors could be separately identified on the basis of these first-order factors. As a result, it was possible to examine either trait correlation after method effects were removed or method correlation after trait effects were accounted for. Table 4 presents the goodness-of-fit indexes of the three second-order CFA models tested.

Model A specified three correlated second-order method factors only, whereas Model B postulated three correlated trait factors only. If any of these models demonstrated acceptable fit, it would mean that the covariation among the nine trait-method combination factors was primarily created by either the trait or method effects only. In evaluating the fit of HCFA models, relying solely on the usual goodness-of-fit indexes such as NNFI and CFI could be misleading. They only indicate the overall ability of models for depicting the indicator variance. In cases where first-order factors are relatively uncorrelated, values of NNFI or CFI of second-order models can be quite high even when the second-order factors explain little of the first-order factor variance. Marsh and Hocevar (1985) proposed an index that is sensitive to the degree of first-order factor correlation that can be used in determining the fit of higher-order models. The target coefficient (TC) roughly represents the proportion of first-order factor variance that is accounted for by the second-order factors. When used along with traditional fit indexes, it provides more accurate information to researchers about the usefulness of the second-order factor structures.

As shown in Table 4, Model A with second-order method factors only did not fit the data well (NNFI = .886, CFI = .893, TC = .644). Model B with second-order trait factors only produced suitable NNFI (.904) and CFI (.911) values, indicating that it was able to account for covariation among the MVs to a reasonable degree. However, it was not able to illustrate the first-order factor covariance to a sufficient degree as evidenced by the TC value of .796. Model C with both trait and method second-order factors not only demonstrated superior fit compared to Model A, $\Delta\chi^2$ (12, $\underline{N}$ = 235) = 466.870, $\underline{p} < .001$, or Model B, $\Delta\chi^2$ (12, $\underline{N}$ = 235) = 264.365, $\underline{p} < .001$, but also displayed an excellent target coefficient (.994). Most of the variance in the first-order factors was thus accounted for by the hypothesized trait and method second-order factors. A second-order model with separate trait and method factors is useful for probing the validity issue because it partitions first-order factor variance into trait, method, and residual (uniqueness) variance components (Marsh & Hocevar, 1988).

Insert Figure 1 about here

Figure 1 shows standardized path coefficients and residual variance of Model C. Evidence of convergent validity can be found in the substantial loadings of first-order factors on their respective second-order trait factors. However, the strong method effects of Problem-Referencing and Task-Referencing factors on their respective lower-order factors should qualify this finding. In comparison, trait effects mostly determined variance of the MSLQ first-order factors. It is also interesting to note that magnitude of the higher-order trait effects were very similar on the Problems and Tasks Self-Efficacy first-order factors (.618 and .608 in

Korean, .675 and .677 in English, and .665 and .718 in math) but noticeably greater on the MSLQ first-order factors (1.0 in Korean, .843 in English, and .848 in math). Problems and Tasks Self-Efficacy factors correlated higher with each other than with MSLQ Self-Efficacy factors. In fact, after the trait effects were removed, the MSLQ method shared virtually no variance with the Problem-Referencing method. This might have caused the disturbance terms of the MSLQ Korean and English Self-Efficacy factors to be fixed to zero (see Figure 1).

Insert Table 6 about here

Table 6 reports correlation coefficients among the second-order factors of Model C. These coefficients are especially helpful in answering the convergent and discriminant validity questions. For example, the correlation coefficients among the second-order self-efficacy factors were obtained after the method effects were accounted for. This allows us to put more faith in the answers generated from these coefficients than those from the correlation among trait-method combination first-order factors. Korean, English, and Math Self-Efficacy factors appeared sufficiently distinct from each other. Correlation coefficients were of moderate value, ranging from .394 to .485. Because these coefficients were not too high and substantially different from unity, discriminant validity of these self-efficacy factors was supported. Deciding whether a given correlation coefficient between any two traits is too high to judge them different calls for a rather subjective judgment (see, e.g., Marsh & Hocevar, 1988, for related discussion). In the present context, these correlation coefficients should be perceived as true trait correlation discussed by Campbell and Fiske (1959) because they represent relations among trait factors that are corrected for unreliability and independent of the shared method variance. The fact that consistent trait correlation was observed across methods in the first-order CFA (Table 5) also supports true trait correlation. Among the second-order method factors, the Problem-Referencing and Task-Referencing factors were most highly correlated (.704). The Task-Referencing factor was also moderately correlated with the MSLQ factor (.460). As mentioned earlier, there was no correlation between the Problem-Referencing and MSLQ factors (.043).

Discussion

The present investigation compared different measures of academic self-efficacy beliefs across varied subject areas and samples, using CFA approaches to the MTMM data. Convergent validity of self-efficacy scores assessed by different methods and discriminant validity of self-efficacy beliefs across multiple academic domains was examined. Results of the first-order CFAs from Studies 1 and 2 showed strong convergence, demonstrating the generalizability of findings. Models that excluded either the trait or the method effects were not able to illustrate the self-efficacy data effectively. The need for different self-efficacy factors confirms the context-specificity of self-efficacy perceptions. When gauging their academic confidence, students responded differently depending on what subject matter area was being tapped by each question. At the same time, the need for different method factors indicates that their responses differed to some degree depending on how the questions were posed and what kind of questions were asked.

When the Campbell and Fiske (1959) guidelines were applied to the first-order CFA models, results generally support the convergent validity of self-efficacy beliefs. Across Studies 1 and 2, students' self-efficacy responses in the same domain assessed by different methods were more highly correlated than self-efficacy scores in different domains assessed by either the same or different methods. Results also verify discriminant validity of self-efficacy responses in

different school subjects. Regardless of how they were assessed, self-efficacy scores in different academic domains did not correlate too highly to cast doubt on their distinctiveness. Most correlation coefficients among self-efficacy factors within the same method were considerably less than unity, despite the fact that they were corrected for attenuation due to measurement errors and hence tended to be higher than what we usually observe in the literature. Even when they were highly correlated as were algebra and geometry self-efficacy scores in Study 1, treating them as products of a single self-efficacy construct resulted in substantial decrement in the model's ability to account for the self-efficacy matrix. Given these results, the moderate correlation among self-efficacy responses should be viewed as evidence of "true" trait correlation rather than lack of discriminant validity (Campbell & Fiske, 1959).

In fact, the moderate correlation among different academic self-efficacy factors found in this study is precisely what the self-efficacy theory would predict (Bandura, 1997). Prior successes and failures in a given domain are the major determinants of people's self-efficacy perceptions in that very domain. These perceptions do generalize, however, to the extent that individuals realize that different domains require similar subskills, dissimilar skills in different domains are acquired and developed concurrently, or success in various domains depends on common self-regulatory capabilities. It is also the case that students' achievement levels in diverse subject areas are often highly correlated. Therefore, it is only reasonable to expect students' self-efficacy judgments in multiple academic subjects to be moderately correlated. The strength of confidence to perform successfully in algebra would be more highly correlated with confidence beliefs in geometry than confidence in, for example, English. However, because there are also skills and competence that are unique to algebra or geometry, strengths of confidence students express toward these two math domains would not be completely identical. Results from the current investigation are consistent with Bong's previous observations with a larger US sample (1997b, $\underline{N}$ = 588) and Korean middle and high school samples (in press-a).

Although results discussed up to this point are coherent across Studies 1 and 2, consistent with previous reports, and reasonable in light of the self-efficacy theory, it should be reminded that they were based on the factors that did not separate out trait effects from method effects or vice versa. In essence, the only major advantage of these first-order CFAs to the traditional zero-order correlation approach is that they account for unreliability in the measures and thus provide more accurate information. They still suffer the same criticisms that the zero-order approach faces. For example, Marsh and Hocevar (1983) wrote, while analyzing the MTMM matrix of nine traits and two methods, "Testing the second and third criteria [proposed by Campbell and Fiske] alone requires that each of the nine convergent validities be compared with 32 different correlations – a total of 288 comparisons. Besides being unwieldy, the likelihood of obtaining rejections due to sampling fluctuations alone increases geometrically with the number of traits and methods" (p. 233). By applying HCFAs, researchers are exempt from making the numerous comparisons they otherwise have to make with the zero-order or the first-order CFA correlation. Yet the HCFAs provide more definitive answers regarding the convergent and discriminant validity of scores by isolating the observed variability among responses into different sources. In Study 2, we were able to partition the variance in each self-efficacy scale into trait, method, and residual variance by performing second-order CFAs. This approach allowed us to examine the degree to which self-efficacy beliefs in different academic subjects correlate, considering the method effects that were apparently in operation. Similarly, it permitted us to assess the degree to which different methods converged with each other, after the trait effects were accounted for.

Results corroborated findings from the first-order models regarding the discriminant validity of various self-efficacy judgments. After the variance due to methods and uniquenesses

were taken out, correlation among self-efficacy beliefs in three subject areas observably decreased and was only moderate in magnitude. When it comes to convergent validity, however, some interesting and also somewhat unsettling results emerged. Students' self-efficacy judgments as assessed by specific problems converged with those estimated using specific task descriptions. On one hand, this may be fully expected because task descriptions in the current study were initially developed from the self-efficacy assessment problems. On the other hand, it gives researchers some reassurance that resorting to generic task descriptions in lieu of particularized problems would produce approximately similar results. For example, investigators have been using task descriptions that are analogous to the ones used in the present study in domains where problem-specific measurement of self-efficacy is not feasible (e.g., Joo et al., 2000; Pajares et al., 1999; Schunk & Ertmer, 1999; Shell et al., 1995). Present findings provide empirical justification for such practice by establishing reasonable equivalence of problem-referenced and task-referenced self-efficacy ratings.

Self-efficacy responses generated by the task-referencing method were also moderately correlated with those from the MSLQ items. However, the strength of their relationship was not to the extent that one could view the two methods comparable. Further, after the trait effects were controlled for, students' self-judged efficacy toward specific problems displayed practically no relationship with self-efficacy ratings from the MSLQ scale. The MSLQ self-efficacy scale differs from the other two measures in its generality. One may argue, therefore, that any observed difference between the MSLQ and problem- or task-specific self-efficacy responses is due to their difference in measurement specificity. This is indeed a plausible assumption according to the self-efficacy theory. Bandura (1997) and other self-efficacy researchers (Pajares, 1996; Zimmerman, 1995) discussed that self-efficacy can and should be assessed at different levels of generality depending on the outcomes of interest. Therefore, one could measure self-efficacy for either performing a particular task under a very specific set of conditions (most specific levels), completing a class of activities sharing the common conditions and properties within the same domain (intermediate levels of specificity), or functioning successfully in given domains without identifying the tasks and conditions under which these tasks are to be performed (most general levels). The problem- and task-referencing methods are akin to the most specific and intermediate levels of specificity, whereas the MSLQ scale resembles the most general levels of self-efficacy assessment. Previous research has demonstrated that correlations among self-efficacy scales within the same domain fluctuate by the measurement specificity (Bong, in press-b). Perceived efficacy measures of different specificity are also proven to be useful for predicting different outcomes (Bong, 1997a; Pajares & Miller, 1995). The present findings may be taken to substantiate these previous observations that evaluation of one's competence generates related but unequal inferences depending on the specificity of the contexts provided.

It is intriguing to note that the variability in students' responses to problem- and task-specific self-efficacy items were determined concomitantly by the type of problems/tasks presented and the subject area from which these problems/tasks were selected. In contrast, students appeared to have provided more or less uniform responses to all MSLQ self-efficacy items as long as they referred to the same academic domain. As a result, method effects of the MSLQ were considerably less than those of problem- and task-referencing methods, once variance due to the subject-specific self-efficacy perceptions was accounted for. Again, the key difference between the problem- and task-referencing methods and the MSLQ scale was whether or not concrete anchors were provided against which to judge perceived self-efficacy. When only general self-efficacy statements were furnished with no explicit reference to particular tasks or conditions, respondents did not distinguish much between these items. It is also worth

mentioning that, perhaps for this reason, students' ratings of their perceived efficacy toward different school subjects were more discrepant in the MSLQ responses compared with their problem- and task-specific judgments in these academic areas. Whether this response tendency is desirable (e.g., internal consistency) or undesirable (e.g., response bias) is an issue that requires further probing.

At minimum, the present results highlight the fact that what and how questions are asked makes a difference in self-efficacy assessment. Researchers would be well advised to pay particular attention to the specificity of self-efficacy measures in view of their predictive and explanatory goals. When devising task-specific self-efficacy measures, they should take heed to what problems and activities to include, given the strong method effects associated with problem- and task-referencing methods. The current investigation also confirmed the effectiveness of higher-order factor analytic procedures for analyzing the convergent and discriminant validity issues. As demonstrated here, results may not show the whole picture and subsequent conclusions could be deficient or misleading when one only looks at the zero-order or first-order factor correlation according to the Campbell-Fiske criteria. We echo with Marsh and Hocevar's (1988) suggestion that HCFA should be the choice of analysis when researchers are dealing with MTMM matrices involving at least three traits and three methods.

The present study yielded some useful findings regarding the equivalence of different self-efficacy measures. Even so, it is one thing to establish convergent validity of measures and another to establish their predictive utility, for people build not only specific task beliefs but also general beliefs that are more than sum of their specific beliefs (Bong & Clark, 1999). The inability of the MSLQ self-efficacy scale to converge with more specific self-efficacy measures may not necessarily mean that the MSLQ taps something other than perceived self-efficacy. Students' responses to the MSLQ self-efficacy items shared a considerable amount of variance with both problem- and task-specific scales when the trait effects remained intact. Nevertheless, after the trait effects were removed, the MSLQ responses had nothing in common with the problem-specific responses. Again, we speculate that difference in the assessment specificity likely have generated such results. However, generality of self-efficacy assessment scales need not be associated with lack of concrete and explicit anchors, which often form the basis of requisite judgments. Asking students to gauge their confidence for obtaining diverse letter grades in the given course (Zimmerman & Bandura, 1994), for example, is a general-level self-efficacy measurement that offers specific anchors. Future research should determine whether and how general self-efficacy perceptions, formed in either the presence or the absence of concrete anchors, differ from more specific self-efficacy perceptions. Provided that differences between the scales observed in this study were mainly a consequence of different measurement specificity, there is little doubt that they would all meaningfully relate to some domain-related outcomes. The question ultimately boils down to selecting a measure that best captures individual's beliefs as they are faced with specific tasks and contingencies (Pajares, 1996). We recommend that future research that aims to evaluate various self-efficacy measures should test predictive validity of those measures, in addition to applying higher-order confirmatory factor analysis.

References

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84, 191-215.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. Child Development, 67, 1206-1222.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. Journal of Personality and Social Psychology, 41, 586-598.

Bentler, P. M. (1992). EQS: Structural equations program manual. Los Angeles, CA: BMDP.

Bong, M. (1997a, March). Congruence of measurement specificity on relations between academic self-efficacy, effort, and achievement indexes. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Bong, M. (1997b). Generality of academic self-efficacy judgments: Evidence of hierarchical relations. Journal of Educational Psychology, 89, 696-709.

Bong, M. (2001). Predictive utility of subject-, task-, and problem-specific self-efficacy judgments for immediate and delayed academic performances. Manuscript submitted for publication.

Bong, M. (in press-a). Between- and within-domain relations of academic motivation among middle and high school students: self-efficacy, task-value, and achievement goals. Journal of Educational Psychology.

Bong, M. (in press-b). Role of self-efficacy and task-value in predicting college students' course performance and future enrollment intentions. Contemporary Educational Psychology.

Bong, M., & Clark, R. E. (1999). Comparison between self-concept and self-efficacy in academic motivation research. Educational Psychologist, 34, 139-154.

Brownstein, S. C., Weiner, M., & Green, S. W. (1994). How to prepare for SAT I. Hauppauge, NY: Barron's Educational Series.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

College Entrance Examination Board and Educational Testing Service (1994). The official guide to SAT II: Subject tests. New York: The College Board.

Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. Educational and Psychological Measurement, 58, 898-915.

Joo, Y. J., Bong, M., & Choi, H. J. (2000). Self-efficacy for self-regulated learning, academic self-efficacy, and Internet self-efficacy in Web-based instruction. Educational Technology Research and Development, 48(2), 5-18.

Kline, R. B. (1998). Principles and practices of structural equation modeling. Guilford Press.

Lent, R. W., Brown, S. D., & Larkin, K. C. (1984). Relation of self-efficacy expectations to academic achievement and persistence. Journal of Counseling Psychology, 31, 356-362.

Marsh, H. W. (1993). Multitrait-multimethod analyses: Inferring each trait-method combination with multiple indicators. Applied Measurement in Education, 6, 49-81.

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. Multivariate Behavioral Research, 33, 181-220.

Marsh, H. W., & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. Journal of Educational Measurement, 20, 231-248.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. Psychological Bulletin, 97, 562-582.

Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. Journal of Applied Psychology, 73, 107-117.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. Journal of Counseling Psychology, 38, 30-38.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. Review of Educational Research, 66, 543-578.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. Journal of Educational Psychology, 86, 193-203.

Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. Journal of Counseling Psychology, 42, 190-198.

Pajares, F., Miller, M. D., & Johnson, M. J. (1999). Gender differences in writing self-beliefs of elementary school students. Journal of Educational Psychology, 91, 50-61.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of Educational Psychology, 82, 33-40.

Roeser, R. W., Midgley, C., Urdan, T. C. (1996). Perceptions of the school psychological environment and early adolescents' psychological and behavioral functioning in school: The mediating role of goals and belonging. Journal of Educational Psychology, 88, 408-422.

Schunk, D. H. (1982). Effects of effort attributional feedback on children's perceived self-efficacy and achievement. Journal of Educational Psychology, 74, 548-556.

Schunk, D. H. (1983). Ability versus effort attributional feedback: Differential effects on self-efficacy and achievement. Journal of Educational Psychology, 75, 848-856.

Schunk, D. H. (1991). Self-efficacy and academic motivation. Educational Psychologist, 26, 207-231.

Schunk, D. H., & Cox, P. (1986). Strategy training and attributional feedback with learning disabled students. Journal of Educational Psychology, 78, 201-209.

Schunk, D. H., & Ertmer, P. A. (1999). Self-regulatory processes during computer skill acquisition: Goal and self-evaluative influences. Journal of Educational Psychology, 91, 251-260.

Schunk, D. H., & Gunn, T. P. (1986). Self-efficacy and skill development: Influence of task strategies and attributions. Journal of Educational Research, 79, 238-244.

Schunk, D. H., & Hanson, A. R. (1985). Peer models: Influence on children's self-efficacy and achievement. Journal of Educational Psychology, 77, 313-322.

Schunk, D. H., & Hanson, A. R. (1989). Self-modeling and children's cognitive skill learning. Journal of Educational Psychology, 81, 155-163.

Schunk, D. H., Hanson, A. R., & Cox, P. (1987). Peer-model attributes and children's achievement behaviors. Journal of Educational Psychology, 79, 54-61.

Shell, D. F., Colvin, C., & Bruning, R. H. (1995). Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. Journal of Educational Psychology, 87, 386-398.

Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), Self-efficacy in changing societies (pp. 202-231). New York: Cambridge University Press.

Zimmerman, B. J. (1996, April). Misconceptions, problems, and dimensions in measuring self-efficacy. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. American Educational Research Journal, 31, 845-862.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. American Educational Research Journal, 29, 663-676.

Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. Journal of Educational Psychology, 91, 241-250.

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. Journal of Educational Psychology, 82, 51-59.

Footnote

[1] The term "trait" is a misnomer for the study of self-efficacy. Bandura (1997) as well as many self-efficacy researchers made it clear that self-efficacy is a context-specific judgment that should not be viewed as one of the personality traits. We decided to retain this term simply to avoid any conceptual difficulty that may arise from using a different term for the well-established multi-trait multi-method procedures.

Table 1

Descriptive Statistics of Scales

Scale	No. Items	Item M	Item SD	α
Study 1				
Problems				
English	7	74.713	21.982	.850
Spanish	7	71.173	32.250	.963
History	7	67.981	24.348	.905
Algebra	7	66.441	27.535	.915
Geometry	7	64.505	28.162	.924
Chemistry	7	55.279	28.656	.891
MSLQ				
English	6	5.512	1.369	.891
Spanish	6	4.850	1.804	.951
History	6	5.393	1.366	.926
Algebra	6	4.829	1.627	.953
Geometry	6	4.473	1.663	.962
Chemistry	6	4.174	1.643	.966
Study 2				
Problems				
Korean	15	69.688	16.583	.957
English	25	65.550	17.005	.974
Math	25	52.999	21.421	.981
Tasks				
Korean	10	61.944	16.207	.937
English	10	59.293	17.269	.962
Math	10	61.068	17.709	.912
MSLQ				
Korean	5	3.154	.760	.872
English	5	3.096	.804	.910
Math	5	3.152	.802	.910

Table 2

Goodness-of-Fit Indexes of Confirmatory Factor Analysis Models Tested in Study 1

Model	Description	χ^2	df	NNFI	CFI	Res.
Null	36 uncorrelated first-order factors	15062.164	630	--	--	--
1	2 correlated method factors only	8441.296	593	.422	.456	.093
2	6 correlated trait factors only	4945.826	579	.671	.697	.092
3	6 correlated trait and 2 correlated method factors with no trait-method correlation	2126.529	542	.872	.890	.058
4	12 correlated trait-method combination factors	1090.211	528	.954	.961	.023
5	Problems Algebra and Problems Geometry factors in Model 4 are combined into Problems Math	1247.941	539	.943	.951	.024
6	MSLQ Algebra and MSLQ Geometry factors are combined into MSLQ Math	1622.179	539	.912	.925	.025
7	Both Problems and MSLQ Algebra and Geometry factors are combined into Problems Math and MSLQ Math	1766.731	550	.903	.916	.027

Note. NNFI = Bentler-Bonnett nonnormed fit index; CFI = comparative fit index; res. = average absolute standardized residuals.

Table 3

Factor Correlations of Model 4 in Study 1

Factors	1	2	3	4	5	6	7	8	9	10	11	12
1. English Problem SE	1.000											
2. Spanish Problem SE	.085	1.000										
3. History Problem SE	.612	.243	1.000									
4. Algebra Problem SE	.520	.106	.419	1.000								
5. Geometry Problem SE	.482	.114	.463	.881	1.000							
6. Chemistry Problem SE	.535	.195	.699	.714	.671	1.000						
7. English MSLQ SE	**.498**	-.065	.319	.187	.102	.217	1.000					
8. Spanish MSLQ SE	.002	**.726**	.117	.080	.068	.136	.177	1.000				
9. History MSLQ SE	.374	-.009	**.475**	.215	.186	.289	.622	.127	1.000			
10. Algebra MSLQ SE	.284	.124	.224	**.644**	.581	.422	.264	.318	.364	1.000		
11. Geometry MSLQ SE	.281	.142	.280	.611	**.663**	.505	.266	.304	.338	.837	1.000	
12. Chemistry MSLQ SE	.348	.163	.391	.520	.497	**.659**	.278	.310	.378	.646	.752	1.000

Note. Bold entries represent mono-trait hetero-method coefficients; Underlined entries represent hetero-trait mono-method coefficients.

Table 4

Goodness-of-Fit Indexes of Confirmatory Factor Analysis Models Tested in Study 2

Model	Description	χ^2	df	NNFI	CFI	Res.	TC
Null	39 uncorrelated first-order factors	12367.610	741	--	--	--	--
1	3 correlated method first-order factors only	6285.800	699	.491	.519	.102	--
2	3 correlated trait first-order factors only	4213.806	699	.680	.698	.084	--
3	3 correlated trait and 3 correlated method first-order factors with no trait-method correlation	1909.269	657	.879	.892	.097	--
4	9 correlated trait-method combination first-order factors	1453.966	666	.925	.932	.035	1.000
4a	9 uncorrelated trait-method combination first-order factors	2787.609	702	.811	.821	.343	0
A	3 correlated method second-order factors	1928.771	690	.886	.893	.077	.644
B	3 correlated trait second-order factors	1726.266	690	.904	.911	.060	.796
C	3 correlated trait and 3 correlated method second-order factors with no trait-method correlation	1461.901	678	.926	.933	.038	.994

Note. NNFI = Bentler-Bonnett nonnormed fit index; CFI = comparative fit index; res. = average absolute standardized residuals; TC = target coefficient ([χ^2 for Model 4a - χ^2 for the model being tested]/[χ^2 for Model 4a - χ^2 for Model 4]).

Table 5

First-Order Factor Correlations of Model 4 in Study 2

Factors	1	2	3	4	5	6	7	8	9
1. Korean Problem SE	1.000								
2. English Problem SE	.638	1.000							
3. Math Problem SE	.506	.537	1.000						
4. Korean Task SE	**.680**	.571	.435	1.000					
5. English Task SE	.460	**.790**	.442	.707	1.000				
6. Math Task SE	.406	.500	**.712**	.587	.582	1.000			
7. Korean MSLQ SE	**.599**	.297	.240	**.580**	.277	.229	1.000		
8. English MSLQ SE	.255	**.588**	.243	.458	**.742**	.379	.415	1.000	
9. Math MSLQ SE	.211	.232	**.555**	.270	.288	**.647**	.369	.403	1.000

Note. Bold entries represent mono-trait hetero-method coefficients; Underlined entries represent hetero-trait mono-method coefficients.

Table 6

Second-Order Factor Correlations of Model C in Study 2

Factors	1	2	3	4	5	6
1. Korean SE	1.000					
2. English SE	.485	1.000				
3. Math SE	.394	.408	1.000			
4. Problem-Referenced	0	0	0	1.000		
5. Task-Referenced	0	0	0	.704	1.000	
6. MSLQ	0	0	0	.043	.460	1.000

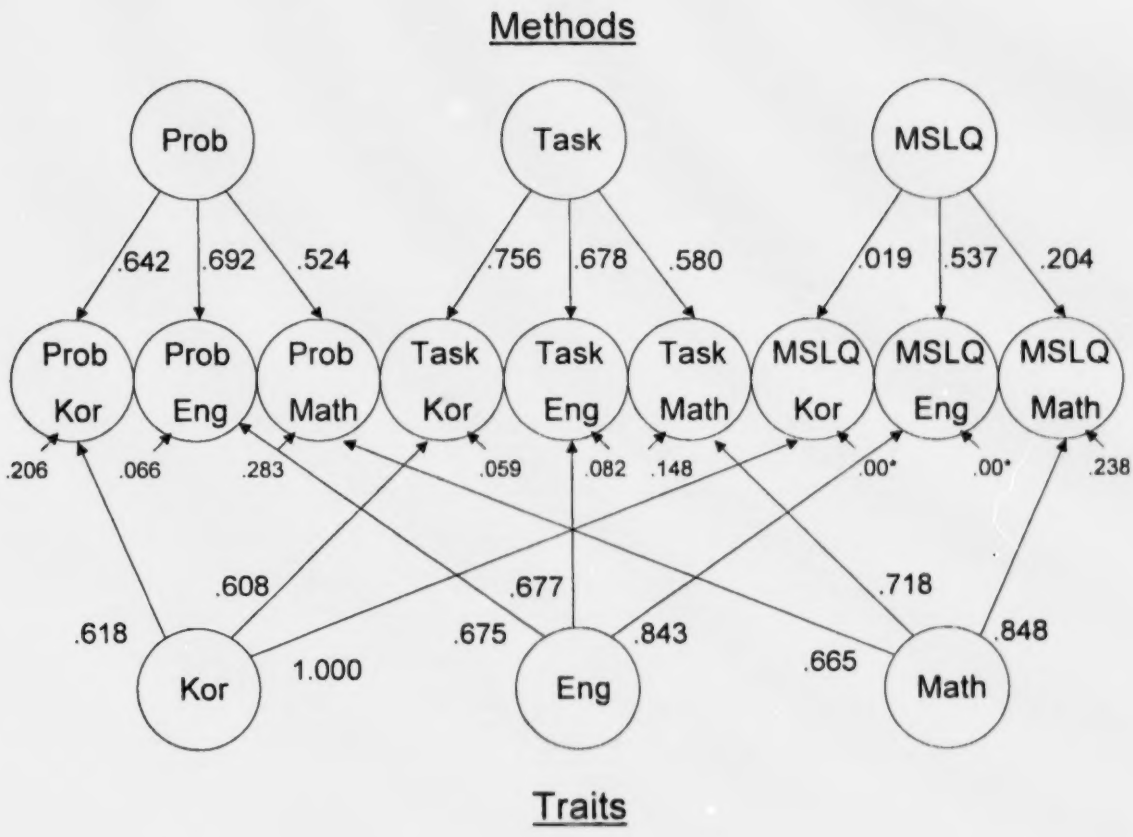

Figure 1. Higher-order confirmatory factor analysis model of multi-trait multi-method data with three correlated trait and three correlated method second-order factors with no trait-method correlation (Model C, Study 2). Prob = Problem-Referenced; Task = Task-Referenced; MSLQ = Motivated Strategies for Learning Questionnaire; Kor = Korean Self-Efficacy; Eng = English Self-Efficacy; Math = Math Self-Efficacy.
* Constrained to be 0.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

TM033736

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title:	Measuring academic self-efficacy: Multi-trait multi-method comparison of scales
Author(s):	MIMI BONG & DENIS HOCEVAR
Corporate Source:	Publication Date: AERA 2001

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ☑	Level 2A ☐	Level 2B ☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Mimi Bong	Printed Name/Position/Title: Associate Prof.
Organization/Address: Mimi Bong, Ph.D Dept of Ed Psychology Univ of South Carolina Columbia SC 29208	Telephone: (803)777-4362 FAX: (803)777-7141 E-Mail Address: mimibong@sc.edu Date: 3/12/2002

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org

EFF-088 (Rev. 2/2001)